# **参考网站：**[**https://rr.peercommunityin.org/**](https://rr.peercommunityin.org/)

Psychological method 最新发表的ICC相关文章,在群里

# 

# 

# **Sections of a Stage 1 Registered Report**

# ***Scientific Reports***

[Abstract 2](#_Toc103777221)

[Introduction 2](#_Toc103777222)

[Methods 4](#_Toc103777223)

[Ethics information 4](#_Toc103777224)

[Secondary Data Description 4](#_Toc103777225)

[Data Collection Procedures 5](#_Toc103777226)

[Experimental design 5](#_Toc103777227)

[Measured Variables 5](#_Toc103777228)

[Stimuli and materials 5](#_Toc103777229)

[Procedure 7](#_Toc103777230)

[Pilot data 7](#_Toc103777231)

[Analysis Plan 8](#_Toc103777232)

[Data availability 10](#_Toc103777233)

[Code availability 11](#_Toc103777234)

[Results 11](#_Toc103777235)

[Discussion 11](#_Toc103777236)

[Acknowledgements 11](#_Toc103777237)

[Author contributions 11](#_Toc103777238)

[Figure Legends 12](#_Toc103777239)

[Table 1. Design Table 12](#_Toc103777240)

[Supplementary information 13](#_Toc103777241)

[References 13](#_Toc103777242)

**Estimating Reliability of the Self-Associative Learning Task as a Measure of Self-Prioritization Effect: Re-analyses of a Longitudinal Dataset**

Zheng Liu 1, Yuanrui Zheng2, Jie Sui3\*, Hu Chuan-Peng 1\*

1 School of Psychology, Nanjing Normal University, Nanjing, China

2 School of Education, Kunming City College, Kunming, China

3 School of Psychology, University of Aberdeen, Old Aberdeen, Scotland

\* Corresponding authors: Hu Chuan-Peng ([hu.chuan-peng@nnu.edu.cn](mailto:hu.chuan-peng@nnu.edu.cn))

# **Abstract**

In recent years, the self-associative learning task (SALT) has provided an effective means to study the regulation of self-related information in individual cognitive processing. However, psychometric properties of the self-associative learning paradigm outcomes have been scarcely reported. Also, in tasks simple as SALT, there are multiple ways to quantify the self-prioritisation effect, such as reaction-time based indices and accuracy-based indices. Thus, it remains unknown (1) whether these indices reliably captures the self-prioritisation effect,and if yes, (2) which indices/index is the most reliable one in group-level and individual-level? In order to fill this gap, we plan to re-analyzed a longitudinal dataset collected in 2016, where 34 healthy volunteers were tested in the self-associative learning task in six sessions (separated by one week). We plan to adopt intraclass correlations and multilevel modelling analysis to achieve an in-depth examination of test-retest reliability as well as practice effect in SALT if one exists. The present study will provide valuable information on SALT for further studies, for example, laying the ground for the future uses of SALT in research, clinical usage, and personal performance monitoring.

# 

# **Introduction**

Self-biases have been shown to be systematic across cognitive domains such as perception, attention, memory, and decision making (Cunningham & Turk, 2017; Desebrock et al., 2018; Sui & Humphreys, 2013). The **self-prioritization effect (SPE)** has long been established as a phenomenon in which people have superior remembrance for information encoded in regard to themselves compared to information encoded in reference to others (Rogers et al., 1977; Symons & Johnson, 1997). People's self-concept is so fundamental that pairing an arbitrary stimulus with it can quickly become salient and influence perception (Humphreys & Sui, 2015; Sui et al., 2015).

Various tests have been administrated to measure the self-prioritization effect, such as the trait-adjectives paradigm (Craik & Tulving, 1975; Rogers et al., 1977), attentional blink paradigm (Shapiro et al., 1997), the ownership task (Cunningham et al., 2008)(see a review on Amodeo et al. (2021)). However, over-learned self-related stimuli, such as the participant's own name or face, are frequently used in research that investigates self-biases in cognition. As a result, it's hard to tell whether such stimuli are processed more quickly because they're associated with the self or because they're more familiar to participants. In recent years, the **self-associative learning task (SALT)** devised by Sui and colleagues has provided an efficient approach to studying the regulation of self-related information in the individual cognitive processing while managed to get around the confound of stimuli familiarity (Sui et al., 2012). The task also rules out the possibility that the SPE is due to a familiarity effect, concreteness, frequency, or length of the words employed. In the task, participants would first learn to associate three geometrical shapes (e.g., triangle, square, and circle) with three different labels (e.g., "You," "friend," and "stranger"). Participants then need to decide whether the following shape-label pairings were appropriately matched or not based on the previously learned relationship in a second stage. Typically, a prominent self-prioritization effect was found with shorter response time, better accuracy and higher sensitivity scores for self-shapes than those of friend and stranger (Schäfer & Frings, 2019; Sel et al., 2019; Sui et al., 2016). The drift-diffusion model was also used to illustrate the perceptual matching task of immediately acquired self-relatedness, with the main effect of self-prioritization being a faster rate of evidence accumulation in the drift-diffusion model – the drift rate, and a bias at the start of the accumulation of information – the starting point (z) (Golubickis et al., 2017; Macrae et al., 2017; Yankouskaya et al., 2020). These effects seem to be robustly replicated in a large number of subsequent studies.

As the SALT is convenient in accessing powerful top-down processing and ability to get around the confound of stimuli familiarity, the use of the associative learning paradigm has grown considerably over the past ten years. For example, because SPE can serve as a trans-diagnostic framework for psychological diseases, clinical and health psychologists have adopted the self-associative learning paradigm to understand atypical self-processing (impaired self-referential cognition) in the clinical population such as autistic population and depressed population (Gillespie‐Smith et al., 2018; Nijhof & Bird, 2019; Sui & Humphreys, 2017). The self-associative learning paradigm is also widely applied in understanding group processes and cross-culture variations (Jiang et al., 2019). It is also modified to be child-friendly to study the automatic self-advantage through child development (Maire et al., 2020).

Surprisingly, although the SALT has been widely applied, there has been microscopic examination and report of the psychometric properties of the self-associative learning paradigm outcomes. To ensure accurate assessment of human perceptual abilities, especially if future study wants to apply the self-associative learning paradigm to clinical settings, such as the diagnosis of depression (Liu et al., 2022), cognitive tests must have high reliability—a high degree of consistency in its measurements (Parsons et al., 2019). Also, in tasks simple as SALT, there are multiple ways to quantify the self-prioritization effect. Thus, it remains unknown (1) whether theses indices reliably capture the self-prioritization effect across times, and if yes, (2) which indices is mostly suitable for repeated measurements?

To fill the current gap, our work seeks to assess the reliability of the self-associative learning task as well as the stability and applicability of the generally used indices of the self-prioritization effect (SPE) in SALT. To this end, we plan to re-analyze a pre-collected dataset in the lab in 2016 where participants were asked to associate three different arbitrary shapes with labels for themselves, a friend, or a stranger across six testing sessions with time intervals for one week. Thus, the current research has three key objectives:

1. Test which index (s) is appropriate and consistent to indicate the group-level self-prioritization effect (SPE) in the SALT;
2. Test which index (s) is consistent to indicate the individual-level self-prioritization effect (SPE) in the SALT;
3. Test whether there is a practice effect across testing sessions.

Our main hypothesis are as follows:

1. (a) Model-based measurements and Reaction time-based measurement are appropriately reliable as group-level SPE indicators in the associative learning task (b) accuracy-based measurement exhibits different degrees of inconsistency from one time point to another.
2. (a) Model-based measurements, which reflect the critical underlying generative process of individuals, are appropriately reliable as individual-level SPE indicators in the associative learning task. (b) RT and accuracy-based measurements exhibits different degrees of inconsistency from one time point to another.
3. (a) There is a practice effect on all indices across testing sessions.

We plan to adopt Hierarchical Linear Model (HLM), Intraclass Correlation Coefficient (ICC) and Split-Half Reliability to test these hypotheses (see Analysis Plan for details). The present study will provide valuable information on SALT for further studies, for example, laying the ground for the prospective uses of SALT in research, clinical usage, and personal performance monitoring.

# **Methods**

## **Ethics information**

As the present research aims to perform a secondary analysis, which does not evolve treatment on humans or animals, informed consent and confidentiality are not an issue in this project. The study where pre-collected dataset was obtained is ethically approved by the research committee at Tsinghua University.

## **Secondary Data Description**

To answer our research questions, we plan to use a pre-collected dataset collected by the Hu Chuan-Peng at Tsinghua University in 2016. The purpose of the original study was to compare the SPE between sub-clinical depressed participants and non-depressed participants. However, the original study only collected the health control group due to the difficult to recruit sub-clinical depressed participants (only 6 participants were collected). The dataset provides six waves (separated by 1 week) of long-term data on 34 non-depressed and 6 depressed participants who were recruited from Tsinghua University community. In each wave, participants completed three parts of tasks: experiment A (a modified SALT), experiment B (a modified SALT) and questionnaires. We plan to **use the subset consisting of the neutral condition in SALT B result of the 34[[1]](#footnote-1) participants with relatively low DBI score**.

## **Data Collection Procedures**

In the study, 36 college students from Tsinghua University community participated the experiment and were compensated. All of them are right-handed and have normal or corrected-to-normal vision. Data of one subject was excluded because of providing confusing participant information to experimenter. Data from one male participant of experiment was missing because of an error in the program. The exclusion left 34 valid participants ( = 21.06, =2.52), with 21 females and 13 males.

## **Experimental design**

Experiment B is a 2 (match vs. not-match) ×3 (id: self, friend, stranger) × 4 (emotion: control, neutral, happy, sad) × 6 (sessions: 1-6) experiment. It is originally designed to compare the self-bias under different emotions (happy, sad, neutral, control).

## **Measured Variables**

In each wave, the participants were recorded on their keypress, the reaction time and accuracy in each trial. The questionaries vary across waves and are related to domains as diverse as personal wellbeing, physical and mental health, and psychological distance between self, friend, stranger.

## **Stimuli and materials**

The experiment was finished individually in a dimly lighted room. Stimuli were presented and responses were collected using E-Prime 2.0 on PC. The monitor was at 1024 × 768 resolution with 100 Hz refresh rate.

The experiment had two phases. Following Sui et al. (2012), the first phase comprised a learning task in which participants were required to associate geometric shapes with labels. The shapes were not presented at this stage. The learning phase lasted for approximately 60 seconds and shape-target associations were counterbalanced across the sample. Next, participants performed a matching task. At the start of each trial, a fixation cross was first displayed in the center of the screen for 500 ms. Then, a shape–label pairing as well as the fixation cross was presented for 100ms, respectively. The next frame showed a blank screen for 1500 ms, or until a response was made. Participants were asked to determine whether the shape was appropriately matched to the label by pressing one of the two response buttons as quickly and precisely as possible within this timeframe.

The participants need to separately learn 4 sets of association between shapes and labels. The associations contain 1 control condition and 3 sets of emotion-based condition. In the control condition, participants learned the association between 3 geometric shapes (circle, horizontal ellipse and vertical ellipse) and three labels (self, friend, stranger). In each of the emotion-based condition, participants would see facial expressions (happy, sad, neutral) appear on the circle, horizontal ellipse and vertical ellipse (see figure 2). In each condition, before beginning the formal experimental trials, participants performed a training session containing 24 practice trials. Following the practice trials, each participant completed 6 blocks of 60 trials in the task. There were six types of shape-label associations: two matches(matched/mismatched) x three shape associations, with 60 trials per association. Participants received a short break (up to 60s) after each block.

Diagram

Description automatically generated

**Figure 2.** Examples of stimuli and time course of the experimental procedure in Experiment B. The labels and feedback appeared in Chinese in the experiment. In the associative learning task, the matched associations of shapes and labels was counterbalanced between participants. Timely feedback was not provided in formal trials.

## **Procedure**

Upon arriving at the laboratory, participants were given written informed consent. After reading and signing the consent, participants finished behavioral experiment A, behavioral experiment B, and questionnaires. The whole experiment was approximately 80 minutes. The participants then completed the same behavioral experiment for five times at the same time in the following five weeks. At each session, participants finished questionnaire: Beck Depression Inventory (BDI)(王振 et al., 2011), Beck Anxiety Inventory (BAI), Positive and Negative Affect Scale (PANAS)(王力 et al., 2007), and state self-esteem scale (Heatherton & Polivy, 1991), as well as psychological distance between self, friend, stranger by visual scale. Also, at the first and last session, participants finish additional questionnaires regarding traits: big-five, Rosenberg Trait self-esteem, IPA (locus of control) (Levenson, 1974; 汪向东 et al., 1999) and belief in free will (Paulhus & Carey, 2010).

## **Pilot data simulated data**

To eliminate possible bias in forming hypothesis, we do not perform any statistical analysis based on the secondary data at stage 1 registration. Instead, we simulated a dataset that has an identical data format as the secondary data. We took an open data of previous research using SALT to examine the self-prioritization effect as the reference to generate the pilot data.

We applied the Bootstrap methods where samples are drawn from the open data of Hu et al. (2020) (accessible at <https://osf.io/mhdsn/>) with replacement (allowing the same sample to appear more than once in the pilot data). Following the format of the secondary data, the pilot data has 6 sessions of 34 participants’ data, where for each participant in each session, there are 24 practice trials and 360 experimental trials (six types of shape-label associations: two matches (matched/mismatched) x three identity associations (self, friend, stranger), with 60 trials per association). Figure 1 shows the first six rows of the pilot data.

Table

Description automatically generated

Figure 1. The first six rows of the pilot data

We ran the pilot data through our proposed statistical analysis to see whether our proposed analysis is appropriate for the secondary data structure (see analysis plan).

## **Analysis Plan**

We will use the Python toolkit HDDM of Bayesian Hierarchical Model (Wiecki et al., 2013) to fit the behavioral data into the drift diffusion model (DDM). All the other analyses will be performed using the statistical software R (R Development Core Team, 2010).

**Data pre-processing**

First, we will pre-process the secondary data using the following criteria (we do not pre-process the secondary data at stage 1 registration):

1. Participant exclusion criteria
2. Participant who has the wrong trial numbers because of procedure errors should be excluded from the analysis.
3. Participants with an overall accuracy < 0.5 should be excluded from the analysis.
4. Participants with any of the conditions with zero accuracy should be excluded from the analysis.
5. Behavioural data exclusion criteria
6. Trials with no response or wrong key press should be excluded from the analysis.
7. Trials with responses less than 200 ms or faster than 1500 ms should be excluded from the analysis.
8. The practice trials will be excluded from the formal analysis.
9. The data under conditions other than the “control condition” will not be used in the current study.

**Calculation of indices & quantifying SPE in the SALT**

Then, we will calculate the indices in the SALT as well as the self-prioritization effect (SPE) revealed by each index at individual-level. We plan to use seven indices which are commonly used in the SALT. In Table 2, we report how the indices is calculated as well as how the self-prioritization effect (SPE) is calculated using these indices.

Table 2. Indices in SALT and corresponding calculation of indices and SPE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Indices ID** | **Indices Calculation** | **SPE Calculation Based on Indices** | | **Source** |
| Mean Reaction times (RT) |  | Type 1 calculation | Self-match - other-match | Sui et al. (2012) |
| Type 2 calculation | self-all - other-all | Sui et al. (2012) |
| Accuracy (ACC) |  | self-match) - other-match | | Sui et al. (2012) |
| d-prime | z-score (ACC (match) - z-score (1 - ACC (non-match)) | self - other | | Sui et al. (2012) |
| Efficiency |  | self-match - other-match | | Stoeber and Eysenck (2008); Sui and Humphreys (2013) |
| Drift rate (v) | DDM：parameters will be identified through model selection | self-match- other-match | | Golubickis et al. (2017) |
| Starting point (z) | self-match - other-match | | Golubickis et al. (2017) |

*Note.* DDM =drift diffusion model.

We will report means and standard deviations of each index in each session, along with other relevant descriptive statistics.

**Reliability of indices in SALT as individual-level/group-level**

The reliability of indices in SALT will be evaluated using intraclass correlations (ICC). The Intraclass correlation coefficient (ICC) is a commonly used reliability statistic in test-retest, intra-rater, and interrater reliability investigations (Fisher, 1970). Compared with Pearson correlation coefficient, the ICC accounts for both degree of correlation and agreement between multiple measurements, making it a more desirable measure of test-retest reliability (Koo & Li, 2016).

Specifically, we will use two-way single-measurement mixed model with absolute agreement between scores of six session (ICC2k) as the reliability measure of group-level SPE across six sessions. For the calculation of ICC2k estimates and their 95% confidence intervals, the formula is:

*Note.* = mean square for rows; = mean square for error; = mean square for columns; = number of subjects; = number of raters/measurements.

We will use a two-way multiple raters random effect model with absolute agreement between scores of six sessions (ICC2 ) as the reliability measure of individual-level SPE across six sessions. For the calculation of ICC2 estimates and their 95% confidence intervals, the formula is:

*Note.* = mean square for rows; = mean square for error; = mean square for columns; = number of subjects.

For ICC2k and ICC values interpretations, we followed the following recommendations: values less than 0.6 are indicative of poor reliability, values between 0.6 and 0.8 indicate substantial reliability, and values greater than 0.8 indicate excellent reliability (Cicchetti & Sparrow, 1981; Kupper & Hafner, 1989).

**Effect related to practice in SALT**

The potential effect of practice will be explored using hierarchical modelling using restricted maximum likelihood estimates with sessions as fixed effects and a random intercept to account for inter-individual differences in baseline performance. Multilevel modelling allows for the comparison of more than two sessions as well as the inclusion of additional predictors such as the number of previous sessions (Ding & Vancleef, 2022). For example, the disparity between results may be smaller with more practice with the test.

We will construct the hierarchical model for each index. The hierarchical model specification was as follows:

Significance will be calculated using the lmerTest package in R(Kuznetsova et al., 2017), which applies Satterthwaite’s method to estimate degrees of freedom and generate *p*-values for mixed models.

A detailed description of correspondence between each hypothesis, each statistical test and interpretation of results is illustrated in the Design Table.

**Split-half reliability of SPE in SALT**

In psychological research, Cronbach’s alpha are often used to calculate the reliability of experiments, however, using Cronbach’s alpha in cognitive experiments often yields biased results. Therefore, more and more studies, use split-half reliability, rather than Cronbach’s alpha, to express the reliability of cognitive experiments. This is because Cronbach’s alpha are calculated based on different experimental conditions, whereas split-half reliabilities are calculated based on experimental trials(Kahveci, 2022). There are four types of split-half reliability: odd-even split-half, front-back split-half, permutation split-half, and Monte Carlo split-half. The odd-even split divided the trials with odd-numbered sequences and even-numbered sequences in half; the first-second split divided the first half of trials and the second half of trials in half; and the permutated split was shuffled the order of trials and randomly assigned one half to one group and the other half to another group. Monte Carlo split-half is similar to permutated split-half. It will repeat thousands of permutated split-half to obtain the average and 95% confidence interval of the split-half reliabilities. This study will mainly adopts Monte Carlo split-half to calculate the split-half reliability of SALT.

First, the data will be stratified according to Session, Match, and Identity. If not stratified, directly spliting the data in half will result in uneven distribution of trials for each experimental condition in the two halves, thereby overestimating or underestimating the reliability of the split. Therefore, after the data is stratified, we split the data. For example, when using Monte Carlo Split-Half, we randomly split the data into two half. Then we repeat this operation 1000 times. This will result in 1000 pairs of two halves of the data. Next, we use these 1000 pairs of data to calculate 1000 Pearson correlation coefficients, and then obtain the average and 95% confidence interval of the Monte Carlo split reliability. As for first-second split, odd-even split, and permutated split, they are similar to Monte Carlo division, but they only perform one split, so only one split-half reliability is obtained without interval estimate of the split-half reliability.

# **Data availability**

We will adhere to the following open science practices: open materials, open data. We will share the raw data, excluding sensitive participants’ information on acceptance of our Stage 2 manuscript. The simulated data is accessible on the Open Science Framework () and GitHub ().

# **Code availability**

Code used to simulate and analyze the pilot data is made accessible in the same location: Open Science Framework () and GitHub ().

# **Results**

**Descriptive Statistics**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Session 6 |
| RT(ms) | 3.96(3.11) | 7.1(31.61) | 3.42(26.81) | -1.67(26.38) | -2.74(21.61) | 4.67(21.78) |
| ACC | 0(.05) | -.01(.06) | -.01(.06) | -.01(.05) | .01(.08) | 0(.06) |
| D-prime | .02(.33) | -.01(.42) | -.04(.25) | -.04(.38) | .06(.39) | .02(.32) |
| Efficiency | 2.79(58.5) | 18.14(75.16) | 1.19(63.51) | 9.66(62.48) | -7.03(85.87) | 9.46(69.07) |
| v(ms) | -57.82(2.65) | -74.95(2.8) | 52.16(2.91) | 37.22(2.55) | -47.73(2.05) | -.19(2.21) |
| z(ms) | 1.12(.67) | 3.63(1.13) | -9.98(.89) | -2.96(.88) | 4.59(.77) | -3.7(.73) |

RT reaction time, ACC accuracy, v drift rate, z starting point

As shown in Table 1, we performed descriptive statistics on the six indicators for each Sessions.

**ICC(Intraclass correlation coefficient)**

Intraclass correlation coefficient (ICC) is a measure of the consistency or reliability of measurements made by different raters (observers) or repeated measurements made by the same rater (observer). In essence, it tells us how much of the variation in the data can be attributed to differences between raters or repeated measurements, and how much of it can be attributed to differences within the subjects being measured. In simple terms, it gives an idea of the proportion of total variation in the data that is due to the true differences between subjects, versus due to measurement error or random fluctuations.

The present study aimed to investigate the stability of six indices, including reaction time (RT), accuracy (ACC), Dprime, Efficiency, drift rate (v) and starting point (z) in the diffusion decision model (DDM), across six time sessions. We use the Intraclass Correlation Coefficients (ICC) to evaluate how much of the variation in SALT can be attributed to within-subject repeatability over time, and how much can be attributed to between-subject differences. Among them, we are most interested in ICC2 and ICC2k, where ICC2 represents the ratio of between-subject variance to total variance, and ICC2k represents the ratio of within-subject variance to total variance. Therefore, we want ICC2 to be as large as possible and ICC2k to be as small as possible, indicating that the differences in our experimental measures are mainly due to between-subject individual differences, and each subject's performance is relatively stable across the six sessions. As shown in Figure 1, the ICC2 values of the six indices are relatively large and ICC2k values are relatively small, supporting our hypothesis.

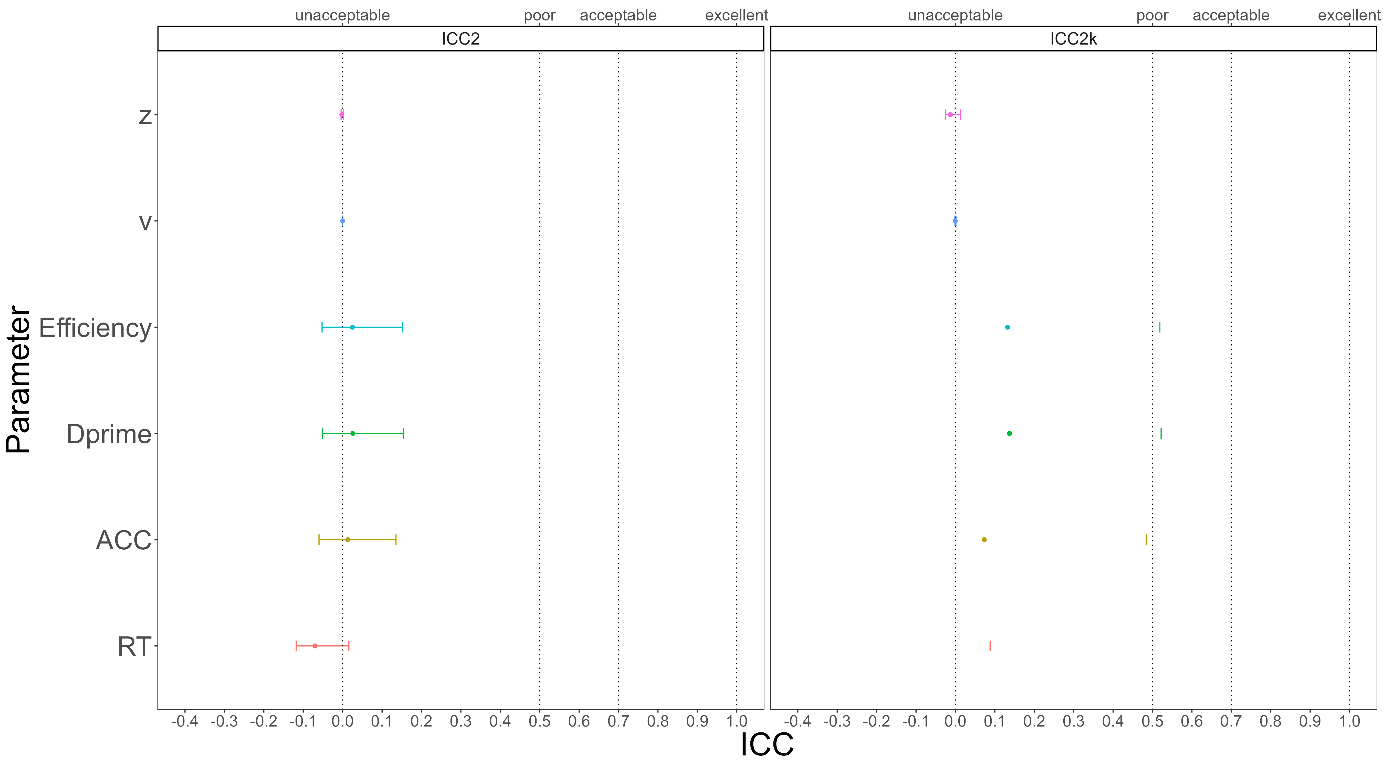


Fig. 1

**Split-Half Reliability**

First, we stratified the data based on three variables: Session, Match, and Identity, and then split the stratified data into two halves using four methods. Next, we calculated the SPE for each of the six indices for each half of the data. Finally, we calculated the split-half reliability for each of the six SPEs. As shown in Figure 2, when using the Monte Carlo split-half, the split-half reliability of the six indices obtained is very high, with the highest value of XXX, which means that it is the most stable of the six SPE indexing calculations for half-confidence. The results from the other three split-half methods were similar to the Monte Carlo method, which will be presented in the supplementary material.

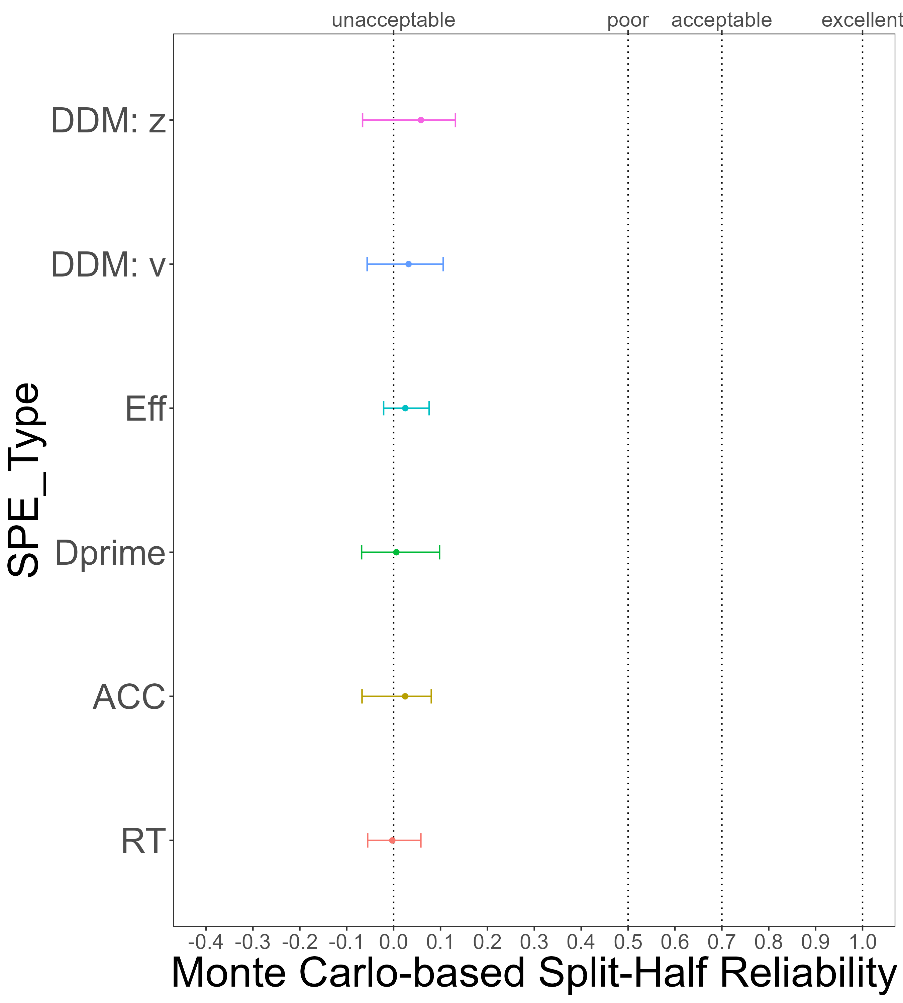


Fig. 2

**HLM(Hierarchical Linear Model)**

The HLM equation for the reaction time (RT\_ms) in the study is as follows: RT\_ms ~ Session\*Match\*Identity + (1|Subject). In this equation, Match and Identity are two independent variables, and Session represents time. In our hypothesis, the results of the SALT experiment should be temporally stable. If the results of the HLM meet our hypothesis, then the variance explained by the Session should be as small as possible, while the variance explained by the Match and Identity should be as large as possible. In the between-subject variance, it should be mainly explained by the Subject. In other words, the conclusion that HLM hopes to get is similar to ICC, and we hope to prove through these two methods that the results of the SALT experiment are temporally stable and that the differences in reaction time are mainly due to individual differences among the subjects.

In our results, the variance between subjects is primarily explained by the subjects themselves, with a regression coefficient of XXX, explaining XXX% of the between-subject variability. The within-subject variance is primarily explained by the experiment variables Match and Identity, with regression coefficients of XXX and XXX respectively, while the regression coefficient of Session is small, XXX. In the HLM results, the reaction time variability is divided into within-subject variability and between-subject variability, with the within-subject variability primarily explained by the two experiment variables Match and Identity. The between-subject variability is primarily explained by the differences between subjects. The variance explained by the subjects themselves is XXX, far greater than the variance explained by Session, XXX. Therefore, the HLM results support our hypothesis that the results of the SALT experiment are stable across time, and that differences in reaction time are primarily due to individual differences and the experiment variables Match and Identity.

# **Discussion**

Do **not** include a **Discussion** section.

# **Acknowledgements**

The present research is support by xxx.

# 

# **Author contributions**

HCP contributed to the conception and supervision of the study and will provide the methodology expertise. JS contributed to fund raising, HCP contributed to data collection. ZL and ZYR will perform the data pre-processing, analysis and visualize the results. In addition, ZL, JS and HCP will contribute to discussing the results and the drafting of the final manuscript. All authors will critically revise the manuscript.

**Competing interests**

The authors declare no competing interests.

**Figures**

You are encouraged to include Figures in the text or at the end of the protocol. Keep in mind that a total of 8 display elements (i.e., combination of Tables and Figures) is permitted in the final, Stage 2, submission. However, to enable typesetting of papers, we advise making the number of display items commensurate with your overall word length (that is, for a shorter paper the number of display items should be lower, for a longer manuscript a higher number may be allowed). Figures/Tables that are not essential should be included in your Supplementary Information file.

# 

# **Figure Legends**

**Figure 1. Guidelines for the preparation of figure captions.** Figure captions should be concise. Begin with a brief title and then describe what is presented in the figure and detail all relevant statistical information. If you show pilot data, list the N of each plot and report full statistics. Aim not to exceed 350 words per legend.

# 

# **Table 1. Design Table**

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Hypothesis (if applicable)** | **Analysis Plan** | **Interpretation given to different outcomes** |
| Which indicator (s) is appropriate and consistent to indicate the group-level self-prioritization effect (SPE) in the SALT? | (a) Model-based measurement (*v, z*) and Reaction time-based measurements (Mean Reaction times) are appropriately reliable as group-level SPE indicators in the associative learning task (b) accuracy-based measurements (accuracy, d-prime, efficiency) exhibits different degrees of inconsistency from one time point to another. | We will use two-way single-measurement mixed model with absolute agreement between scores of six session (ICC2k) as reliability measure of group-level SPE across six sessions. | ICC 2k values less than 0.6 are indicative of poor reliability, values between 0.6 and 0.8 indicate substantial reliability, values greater than 0.8 indicate excellent reliability. |
| Which indicator (s) is appropriate and consistent to indicate the individual-level self-prioritization effect (SPE) in the SALT? | (a) Model-based measurement (*v, z*), which may reflect the critical underlying generative process of individuals, are appropriately reliable as individual-level SPE indicators in the associative learning task. (b) RT and accuracy-based measurements (Mean Reaction times, accuracy, d-prime, efficiency) exhibit different degrees of inconsistency from one time point to another. | We will use a two-way multiple raters random effect model with absolute agreement between scores of six sessions (ICC2) as reliability measure of individual-level SPE across six sessions. | ICC 2 values less than 0.6 are indicative of poor reliability, values between 0.6 and 0.8 indicate substantial reliability, values greater than 0.8 indicate excellent reliability. |
| Is there a practice effect across testing sessions? | There is a practice effect on all indices across testing sessions. | The effect of practice will be explored using hierarchical modelling using restricted maximum likelihood estimates with sessions as fixed effects and a random intercept to account for inter-individual differences in baseline performance. Significance will be calculated using Satterthwaite’s method to estimate degrees of freedom and generate *p-*values for mixed models. | *p*<0.05 as evidence for the presence of a practice effect. |

# **Supplementary information**

Please report pilot data in detail here and include any other material that provides background information.

Supplementary Table 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SPE\_type | SH\_type | SH\_r |  | SPE\_type | SH\_type | SH\_r |
| RT | First-Second | .01 |  | Efficiency | First-Second | .07 |
| RT | Odd-Even | -.05 |  | Efficiency | Odd-Even | -.04 |
| RT | Permuted | .01 |  | Efficiency | Permuted | .05 |
| ACC | First-Second | .02 |  | DDM: v | First-Second | .04 |
| ACC | Odd-Even | -.05 |  | DDM: v | Odd-Even | -.05 |
| ACC | Permuted | .07 |  | DDM: v | Permuted | .10 |
| Dprime | First-Second | .01 |  | DDM: z | First-Second | .07 |
| Dprime | Odd-Even | -.08 |  | DDM: z | Odd-Even | .01 |
| Dprime | Permuted | -.02 |  | DDM: z | Permuted | .13 |

# **References**

Amodeo, L., Wiersema, J. R., Brass, M., & Nijhof, A. D. (2021). A comparison of self-bias measures across cognitive domains. *BMC Psychology, 9*(1), 1-132. <https://doi.org/10.1186/s40359-021-00639-x>

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic, 86*(2), 127-137.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294. <https://doi.org/10.1037/0096-3445.104.3.268>

Cunningham, S. J., & Turk, D. J. (2017). Editorial: A review of self-processing biases in cognition. *Quarterly journal of experimental psychology (2006), 70*(6), 987-995. <https://doi.org/10.1080/17470218.2016.1276609>

Cunningham, S. J., Turk, D. J., Macdonald, L. M., & Macrae, C. N. (2008). *Yours or mine? Ownership and memory* [doi:10.1016/j.concog.2007.04.003]. Elsevier Science.

Desebrock, C., Sui, J., & Spence, C. (2018). Self-reference in action: Arm-movement responses are enhanced in perceptual matching. *Acta psychologica, 190*, 258-266. <https://doi.org/10.1016/j.actpsy.2018.08.009>

Ding, X., & Vancleef, K. (2022). Test–retest reliability and practice effect of the Leuven Perceptual Organisation Screening Test. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01741-z>

Fisher, R. A. (1970). *Statistical methods for research workers* (14th ed., revised and enlarged. ed.). Oliver and Boyd.

Gillespie‐Smith, K., Ballantyne, C., Branigan, H. P., Turk, D. J., & Cunningham, S. J. (2018). The I in autism: Severity and social functioning in autism are related to self‐processing. *British journal of developmental psychology, 36*(1), 127-141. <https://doi.org/10.1111/bjdp.12219>

Golubickis, M., Falben, J. K., Sahraie, A., Visokomogilski, A., Cunningham, W. A., Sui, J., & Macrae, C. N. (2017). Self-prioritization and perceptual matching: The effects of temporal construal. *Mem Cognit, 45*(7), 1223-1239. <https://doi.org/10.3758/s13421-017-0722-3>

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60*(6), 895-910. <https://doi.org/10.1037/0022-3514.60.6.895>

Hu, C.-P., Lan, Y., Macrae, C. N., & Sui, J. (2020). Good Me Bad Me: Prioritization of the Good-Self During Perceptual Decision-Making. *Collabra. Psychology, 6*(1), 20. <https://doi.org/10.1525/collabra.301>

Humphreys, G. W., & Sui, J. (2015). The salient self: Social saliency effects based on self-bias. *Journal of cognitive psychology (Hove, England), 27*(2), 129-140. <https://doi.org/10.1080/20445911.2014.996156>

Jiang, M., Wong, S. K. M., Chung, H. K. S., Sun, Y., Hsiao, J. H., Sui, J., & Humphreys, G. W. (2019). Cultural Orientation of Self-Bias in Perceptual Matching. *Front Psychol, 10*, 1469. <https://doi.org/10.3389/fpsyg.2019.01469>

Kahveci, S., Bathke, A., & Blechert, J. . (2022). Reliability of reaction time tasks: how should it be computed? *Preprint*, 1-30. <https://doi.org/10.31234/osf.io/ta59r>

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine, 15*(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>

Kupper, L. L., & Hafner, K. b. (1989). On Assessing Interrater Agreement for Multiple Attribute Responses. *Biometrics, 45*(3), 957-967. <https://doi.org/10.2307/2531695>

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of statistical software, 82*(13), 1-26. <https://doi.org/10.18637/jss.v082.i13>

Levenson, H. (1974). Activism and Powerful Others: Distinctions within the Concept of Internal-External Control. *Journal of Personality Assessment, 38*(4), 377-383. <https://doi.org/10.1080/00223891.1974.10119988>

Liu, Y. S., Song, Y., Lee, N. A., Bennett, D. M., Button, K. S., Greenshaw, A., Cao, B., & Sui, J. (2022). Depression screening using a non-verbal self-association task: A machine-learning based pilot study. *Journal of Affective Disorders*. <https://doi.org/https://doi.org/10.1016/j.jad.2022.04.122>

Macrae, C. N., Visokomogilski, A., Golubickis, M., Cunningham, W. A., & Sahraie, A. (2017). Self-Relevance Prioritizes Access to Visual Awareness. *Journal of experimental psychology. Human perception and performance, 43*(3), 438-443. <https://doi.org/10.1037/xhp0000361>

Maire, H., Brochard, R., & Zagar, D. (2020). A Developmental Study of the Self‐Prioritization Effect in Children Between 6 and 10 Years of Age. *Child development, 91*(3), 694-704. <https://doi.org/10.1111/cdev.13352>

Nijhof, A. D., & Bird, G. (2019). Self‐processing in individuals with autism spectrum disorder. *Autism research, 12*(11), 1580-1584. <https://doi.org/10.1002/aur.2200>

Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological Science Needs a Standard Practice of Reporting the Reliability of Cognitive-Behavioral Measurements. *Advances in methods and practices in psychological science, 2*(4), 378-395. <https://doi.org/10.1177/2515245919879695>

Paulhus, D. L., & Carey, J. M. (2010). The FAD–Plus: Measuring Lay Beliefs Regarding Free Will and Related Constructs. *Journal of Personality Assessment, 93*(1), 96-104. <https://doi.org/10.1080/00223891.2010.528483>

R Development Core Team. (2010). *R: A language and enviornment for statistical computing.* In R Foundation for Statisticial Computing

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*(9), 677-688. <https://doi.org/10.1037/0022-3514.35.9.677>

Schäfer, S., & Frings, C. (2019). Understanding self-prioritisation: the prioritisation of self-relevant stimuli and its relation to the individual self-esteem. *Journal of cognitive psychology (Hove, England), 31*(8), 813-824. <https://doi.org/10.1080/20445911.2019.1686393>

Sel, A., Sui, J., Shepherd, J., & Humphreys, G. (2019). Self-Association and Attentional Processing Regarding Perceptually Salient Items. *Review of philosophy and psychology, 10*(4), 735-746. <https://doi.org/10.1007/s13164-018-0430-3>

Shapiro, K. L., Caldwell, J., & Sorensen, R. E. (1997). Personal names and the attentional blink: a visual "cocktail party" effect. *J Exp Psychol Hum Percept Perform, 23*(2), 504-514. <https://doi.org/10.1037//0096-1523.23.2.504>

Stoeber, J., & Eysenck, M. W. (2008). Perfectionism and efficiency: Accuracy, response bias, and invested time in proof-reading performance. *Journal of research in personality, 42*(6), 1673-1678. <https://doi.org/10.1016/j.jrp.2008.08.001>

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of experimental psychology. Human perception and performance, 38*(5), 1105-1117. <https://doi.org/10.1037/a0029792>

Sui, J., & Humphreys, G. W. (2013). Self-referential processing is distinct from semantic elaboration: Evidence from long-term memory effects in a patient with amnesia and semantic impairments. *Neuropsychologia, 51*(13), 2663-2673. <https://doi.org/10.1016/j.neuropsychologia.2013.07.025>

Sui, J., & Humphreys, G. W. (2017). The self survives extinction: Self-association biases attention in patients with visual extinction. *Cortex, 95*, 248-256. <https://doi.org/10.1016/j.cortex.2017.08.006>

Sui, J., Liu, M., Mevorach, C., & Humphreys, G. W. (2015). The salient self: the left intraparietal sulcus responds to social as well as perceptual-salience after self-association. *Cereb Cortex, 25*(4), 1060-1068. <https://doi.org/10.1093/cercor/bht302>

Sui, J., Ohrling, E., & Humphreys, G. W. (2016). Negative mood disrupts self- and reward-biases in perceptual matching. *Q J Exp Psychol (Hove), 69*(7), 1438-1448. <https://doi.org/10.1080/17470218.2015.1122069>

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin, 121*(3), 371-394. <https://doi.org/10.1037/0033-2909.121.3.371>

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in neuroinformatics, 7*, 14-14. <https://doi.org/10.3389/fninf.2013.00014>

Yankouskaya, A., Bührle, R., Lugt, E., Stolte, M., & Sui, J. (2020). Intertwining personal and reward relevance: evidence from the drift-diffusion model. *Psychol Res, 84*(1), 32-50. <https://doi.org/10.1007/s00426-018-0979-6>

汪向东, 王希林, & 马弘. (1999). *心理卫生评定量表手册*. 中国心理卫生杂志社.

王力, 李中权, 柳恒超, & 杜卫. (2007). PANAS-X总维度量表在中国人群中的因素结构. *中国临床心理学杂志*(06), 565-568.

王振, 苑成梅, 黄佳, 李则挚, 陈珏, 张海音, 方贻儒, & 肖泽萍. (2011). 贝克抑郁量表第2版中文版在抑郁症患者中的信效度. *中国心理卫生杂志*(06), 476-480.

1. Based on the average effect size of group-level SPE reported by Sui et al. (2012), G\*Power (f = .40, α = .05, power = 80%) revealed a minimal requirement of 16 participants. Thus, the sample size in the secondary data is sufficient to detect the self-prioritization effect at group-level. [↑](#footnote-ref-1)